PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G10L 5/06 // 3/00	A1	(11) International Publication Number: **WO 98/40875** (43) International Publication Date: 17 September 1998 (17.09.98)

(21) International Application Number: PCT/SE98/00432

(22) International Filing Date: 10 March 1998 (10.03.98)

(30) Priority Data:
9700898-1 13 March 1997 (13.03.97) SE

(71) Applicant *(for all designated States except US)*: TELIA AB (publ) [SE/SE]; Mårbackagatan 11, S-123 86 Farsta (SE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MELIN, Håkan [SE/SE]; Gustavavägen 19, S-178 31 Ekerö (SE). SUNDBERG, Erik [SE/SE]; Köpmangatan 4,4, S-111 31 Stockholm (SE).

(74) Agent: PRAGSTEN, Rolf; Telia Research AB, Vitsandsgatan 9, S-123 86 Farsta (SE).

(81) Designated States: EE, JP, LT, LV, NO, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SPEAKER VERIFICATION SYSTEM

(57) Abstract

The invention relates to a method at a speaker verification system/speaker identification system which makes possible for the system operator to find out the identity of a customer by means of analysis of a recording of the customer's speech. In speaker verification systems/speaker identification systems the amount of voice data which are collected from the customer is a decisive limit for the usefulness of the system. The more parameters a model comprises, the better it can be adapted to given training data at the same time as there is needed more and more training data to in a reliable way be able to estimate all existing parameters. The invention utilizes pre-trained reference models in a speaker model so that one can benefit from collected data in addition to the information which the customer himself/herself speaks in his/her registration call. The central concept of the invention is to organize said reference models in a set of pro-models and anti-models, where the pro-models model a quality which the customer has, and the anti-models model a quality which the customer does not have.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

TITLE OF THE INVENTION: SPEAKER VERIFICATION SYSTEM

FIELD OF THE INVENTION

The present invention relates to a method at a speaker verification system/speaker identification system which makes possible for the system operator to find out the identity of a customer by means of analysis of a recording of the customer´s voice data.

PRIOR ART

In speaker verification systems, systems for automatic verification of the identity of a speaker, the amount of voice data which has to be collected from the customer is a decisive limit to the use. The more parameters a model has, the better it can be adapted to given training data, but at the same time there is needed more and more training data (time which the customer has to spend in an initial phase) to in a reliable way estimate all parameters.

A problem in connection with speaker verification consequently is to create a sufficiently good model of a customer´s voice, on the basis of as small an amount of voice data as possible, in order to find out the identity of the customer by means of analysis of the recording of the customer´s voice data. By customer here is meant a user of some service with need of check of authorization.

One consequently might say that the above mentioned problem is a type of optimization problem where it is a matter of utilizing smallest possible amount of voice data to in a reliable way be able to appoint the identity of the speaker.

The aim with the present invention consequently is to solve the above mentioned problem.

SUMMARY OF THE INVENTION

The above mentioned aim is solved by means of a method at a speaker verification system/speaker identification system which is presented in the characterizing part of the patent claim 1.

The present invention has the advantage, in comparison with previous speaker verification systems/speaker identification systems, that, in spite of utilizing a minimal amount of voice data, the identity of the speaker can quickly be found out.

Further characteristics are given in the subclaims.

DETAILED DESCRIPTION OF EMBODIMENTS OF THE INVENTION

In technical connections one usually makes a difference between speaker identification and speaker verification.

With speaker identification then is meant a verification system where a speaker identifies himself/herself by speaking just any sentences, at which the identification system analyses the voice and identifies characteristics of the voice by which the speaker identification is performed.

With speaker verification is meant a verification system where a speaker´s identity is verified by the speaker speaking (or entering by keypad) a specific in advance decided information, at which the verification system directly confirms the authenticity of the information (and identity) or rejects it (example of such a system is a cash dispenser; in Sweden "Bankomat").

The two systems basically relate to the same thing, which is to distinguish and distinctly find out a speaker´s identity.

It consequently should be realized that in the present invention we equate the concepts "speaker verification" and "speaker identification".

The invention is inteded to be used in all speaker verification systems, especially in such which are used in a service where one has access to information about the users.

The voice recording can be made either directly at the equipment, where the verification is performed, or be transferred via different media. Medium can be telephone or other telecommunication media, inclusive computers.

In speaker verification systems today often a "likelihood normalization" is utilized, i.e. a type of probability normalization. In principle these speaker verification systems function in the following way.

Let us suppose that a customer, for instance Leif, has the intention to verify his identity by means of a speaker verification system to get access to a certain service. In this case is assumed that Leif´s voice profile is stored since before in a database belonging to the speaker verification system.

When Leif speaks a voice message via for instance a telephone in the speaker verification unit, the voice profile is stored and analysed. The speaker verification unit finds out that the probability is very high that Leif is a man over 40 years old. In addition the speaker verification unit finds out that Leif is speaking staccato. The speaker verification unit now searches in the hierarchy of different groups in the database and finds a group comprising men over 40 who are speaking staccato.

This group is rather limited (for instance 40 persons) and the speaker verification unit compares Leif´s stored voice profile with all voice profiles which are stored in this special group. With very great probability the speaker verification unit consequently finds Leif´s voice profile in this group, whereupon identification is made.

The above mentioned method consequently is based on that the speaker verification unit on basis of probability finds out to which group in a database a person, for

instance Leif, belongs. After that, stored voice profile is compared with all voice profiles in said group.

This method of course is considerably more efficient than if the speaker verification unit indiscriminately should compare stored voice profile with all included voice profiles in the database. This would take an enormous amount of time if the database for instance contained some thousand voice profiles.

The present invention is a further development and improvement of the above mentioned method and is based on that one by using pre-trained reference models as components in a speaker model can benefit from collected data in addition to these which a customer himself/herself speaks in his/her call to be recorded, and by that reduce the lenght of this call. The central idea of the invention is to organize these reference models in a set of pro-models and anti-models. The idea is that the pro-models shall model a quality which the customer has (for instance woman, beween 20 and 25 years) and the anti-model a quality which the customer does not have (for instance man, not between 20 and 25 years). Purely mathematically produced reference models which normally do not correspond to a distinguishable quality of the customer also can be used.

Complementary sets of pro- and anti-models should be used. If the reference models correspond to concrete qualities of the customer, in addition *a priori* knowledge can be used to control the selection of reference models. This knowledge in different ways can be made accessible in the system.

A more detailed description will be given a bit further on in the description.

In speaker verification connections one uses, as has been mentioned above, "likelihood normalization" where one standardizes the contribution from a customer specific model with one or more "world models" or "impostor models", which with above used terminology are anti-models. The

customer specific model corresponds to function f_c in equation (1) below. The novelty in (1) therefore is to combine the anti-models with (complementary) pro-models. Whether it is a novelty to generally make use of *a priori* knowledge to select reference models is doubtful, but the arrangement with pro- and anti-models goes well with the use of *a priori* knowledge. The theory of selecting an optimal set of reference models and a belonging projection is certainly known in the mathematics/signal theory, and consequently is no novelty in itself, but the application of this thinking in connection with speaker verification is according to our opinion quite a novelty.

In the following the invention will be described in more detail.

Regard a speaker model as consisting of a) reference models, and b) a projection on these reference models. The projection can for instance be a weighted sum of contributions from the reference models (a linear combination). In addition a speaker model of course can include model elements which are built exclusively from voice material from the customer himself/herself and which do not use reference models, but the following description focuses on the part where some form of reference model is included.

The reference models are normally trained from speech in a database which is collected in the design phase of the system, i.e. before a customer registers himself/herself in the system. A reference model either can model I) some predetermined entity (for instance "female speaker", "speaker under 16 years" or "call from GSM-telephone"), or II) something determined by mathematical optimization and which by that cannot very well be connected to a specific *a priori* knowledge as in case I.

Arrange the reference models in one set with pro-models and one set with anti-models and calculate the "hit probability", P, of the total model, so that contributions

from the pro-models increase P, and contributions from the anti-models reduce P. This procedure can mathematically be expressed according to equation (1), where f_p and f_a are functions of contributions of the pro- respective anti-models, and together constitute the projection part of the total model. f_c is a function of submodels trained on data from the customer himself/herself. One also can make use of a logarithmic variant of (1).

$$P = f_c(c_1, c_2, \ldots, c_N) \frac{f_p(p_1, p_2, \ldots\ldots, p_M)}{f_a(a_1, a_2, \ldots\ldots, a_Q)}$$

If reference models according to case I are used, one can utilize *a priori* knowledge about the customer to build the customer´s speaker model, for instance knowledge about the speaker´s sex, by selecting right reference models.

Example: For a male speaker one can select a pro-model for "male speaker" and an anti-model for "female speaker".

In this way one can in a simple way make benefit from *a priori* knowledge when one builds ones speaker model. This knowledge will be a contribution to collected voice data and one can make a better functioning model with less collected voice data from respective customer. One suitably selects complementary reference models as pro- and anti-models as in the example above. In this way one ought to get a more reliable, balanced model and one can get a discrimination effect by the two complementary models "pulling" in different directions.

The above mentioned "*a priori* knowledge" may be introduced into the system at different phases and in different ways:

a) At the (first) call to be recorded and by that in connection with building of the first speaker model. If the customer already is registered in the service and therefore identifies himself/herself to the system at the

registration to the speaker verification system, one can utilize customer information which is already stored in a database, for instance sex and age. If the customer is not pre-registered in the service, he/she may present his/her civic registration number at the registration call, and then one can get information about the sex by looking at the civic registration number. One also explicitly can ask about sex and age during the call.

b) After the first call to be recorded. At that, one may already have taken the model into operation, and it will be a matter of rebuilding the model with new information. The information can for instance come from a filled up and sent in form which the customer signs to be allowed to go on with the service after an initial phase. Adaption of speaker model and especially change of topology during its life cycle is treated in Telia´s patent application No 9602622-4 relating to "Procedure and arrangement for adaption at for instance speaker verification systems" (Case 520).

Instead of using pure *a priori* knowledge, one may select one´s reference models by calculating an optimal set of reference models and belonging projection on these.

The above is only to be regarded as advantageous embodiments of the invention, and the extent of protection of the invention is only defined by what is indicated in the following patent claims.

PATENT CLAIMS

1. Method at a speaker verification system/speaker identification system which makes possible for a system operator to find out a customer's identity by means of analysis of a recording of the customer's voice data, c h a r a c t e r i z e d in that reference models are organized in a set of pro-models and a set of anti-models which constitute components in a speaker model, which speaker model on probability basis is utilized by said speaker verification system/speaker identification system to process and benefit from said collected voice data in addition to the data which the customer himself/herself speaks at his/her call to be recorded, whereby the customer's identity can be found out with minimal recorded voice data.

2. Method according to patent claim 1, c h a r a c t e r i z e d in that said pro-models model qualities which the customer has, and said anti-models model qualities which the customer does not have.

3. Method according to patent claim 1, c h a r a c t e r i z e d in that said reference models are trained to recognize just any voice information, which voice information has been stored in a database at the design phase of said speaker verification system/speaker identification system before a customer has had time to register himself/herself in said system.

4. Method according to any of the previous patent claims, c h a r a c t e r i z e d in that said reference model is arranged in said set of pro-models and said set of anti-models, at which the "hit probability", P, of said reference model is calculated so that contributions from the pro-models increase P, and contributions from the anti-models reduce P.

5. Method according to patent claim 4, c h a r a c t e r i z e d in that the "hit probability" of said total reference model, i.e. total probability that a

certain customer belongs to a certain category, is given by the formula:

$$P = f_c\,(C_1, C_2, \ldots, C_N)\,\frac{f_p\,(P_1, P_2, \ldots\ldots, P_M)}{f_a(a_1, a_2, \ldots\ldots, a_Q)}$$

where P corresponds to the "hit probability" of the reference model, f_p and f_a are functions of contributions of the pro- respective anti-models which together constitute the projection part of the total model, and f_c is a function of submodels trained on voice data from the customer himself/herself.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 98/00432

A. CLASSIFICATION OF SUBJECT MATTER

IPC6: G10L 5/06 // G10L 3/00
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DIALOG, WPIL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0744734 A2 (AT&T IPM CORP.), 13 May 1996 (13.05.96), page 3, line 5 - line 19; page 4, line 16 - line 21, claim 1, abstract	1-2
Y		3
A		4-5
	--	
A	WO 9617341 A1 (THE AUSTRALIAN NATIONAL UNIVERSITY), 6 June 1996 (06.06.96), page 4, line 10 - page 5, line 11, abstract	1-2,4-5
Y		3
	--	

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" erlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 June 1998	10 -07- 1998
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Linus Wretblad Telephone No. +46 8 782 25 00

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 98/00432

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 9505656 A1 (THE UNIVERSITY OF QUEENSLAND), 23 February 1995 (23.02.95), page 3, line 10 - line 15; page 10, line 11 - line 19, abstract --	1-5
A	US 5522012 A (RICHARD J. MAMMONE ET AL), 28 May 1996 (28.05.96), column 2, line 33 - column 3, line 11; column 4, line 5 - line 19, abstract --	1-5
A	US 5598507 A (KIMBER DONALD G ET AL), 28 January 1997 (28.01.97), abstract --	1-3
A	JP 7261785 A (ATR ONSEI HONYAKU TSUSHIN KENKYUSHO:KK), 13 October 1995 (13.10.95), abstract -- --------	1-3

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

09/06/98

International application No.
PCT/SE 98/00432

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0744734 A2	13/05/96	CA 2173302 A	23/11/96
		JP 8314491 A	29/11/96
		US 5687287 A	11/11/97
WO 9617341 A1	06/06/96	AU 683783 B	20/11/97
		AU 4110996 A	19/06/96
		AU PM983094 D	00/00/00
		CA 2206628 A	06/06/96
		EP 0801787 A	22/10/97
WO 9505656 A1	23/02/95	AU 7378694 A	14/03/95
US 5522012 A	28/05/96	AU 683370 B	06/11/97
		AU 2116495 A	11/09/95
		CA 2184256 A	31/08/95
		CN 1142274 A	05/02/97
		EP 0748500 A	18/12/96
		JP 10500781 T	20/01/98
		WO 9523408 A	31/08/95
US 5598507 A	28/01/97	JP 7287593 A	31/10/95
JP 7261785 A	13/10/95	NONE	